



Investigating speciation in face of polyploidization: what can we learn from approximate Bayesian computation approach?

Camille Roux and John Pannell

bioRxiv first posted online February 9, 2014

Access the most recent version at doi: <http://dx.doi.org/10.1101/002527>

**Creative
Commons
License**

The copyright holder for this preprint is the author/funder. It is made available under a [CC-BY 4.0 International license](#).

Investigating speciation in face of polyploidization: what can we learn from approximate Bayesian computation approach?

Authors : Camille Roux¹ and John R. Pannell¹

Affiliations:

¹Department of Ecology and Evolution, Lausanne University, Biophore/Sorge, Lausanne, Switzerland.

Corresponding author address:

Camille Roux

Department of Ecology and Evolution, UNIL, Biophore.

1015-Lausanne, Switzerland.

E-mail: camille.roux.1@unil.ch

phone: +4121 692 4263

Running title: Robustness of ABC to investigate polyploidization.

Keywords: polyploidization, speciation, approximate Bayesian computation, Capsella.

ABSTRACT

Despite its importance in the diversification of many eucaryote clades, particularly plants, detailed genomic analysis of polyploid species is still in its infancy, with published analysis of only a handful of model species to date. Fundamental questions concerning the origin of polyploid lineages (e.g., auto- vs. allopolyploidy) and the extent to which polyploid genomes display disomic vs. polysomic vs. heterosomic inheritance are poorly resolved for most polyploids, not least because they have hitherto required detailed karyotypic analysis or the analysis of allele segregation at multiple loci in pedigrees or artificial crosses, which are often not practical for non-model species. However, the increasing availability of sequence data for non-model species now presents an opportunity to apply established approaches for the evolutionary analysis of genomic data to polyploid species complexes. Here, we ask whether approximate Bayesian computation (ABC), applied to sequence data produced by next-generation sequencing technologies from polyploid taxa, allows correct inference of the evolutionary and demographic history of polyploid lineages and their close relatives. We use simulations to investigate how the number of sampled individuals, the number of surveyed loci and their length affect the accuracy and precision of evolutionary and demographic inferences by ABC, including the mode of polyploidisation, mode of inheritance of polyploid taxa, the relative timing of genome duplication and speciation, and effective populations sizes of contributing lineages. We also apply the ABC

framework we develop to sequence data from diploid and polyploidy species of the plant genus *Capsella*, for which we infer an allopolyploid origin for tetra *C. bursa-pastoris* \approx 90,000 years ago. In general, our results indicate that ABC is a promising and powerful method for uncovering the origin and subsequent evolution of polyploid species.

INTRODUCTION

Our ability to use of contemporary patterns of genetic variation to infer and understand past evolutionary and demographic processes (Harrison 1993) has taken great strides in recent years with the availability of inexpensive sequence data and new statistical tools to analyze it. One of the aims of such projects is to infer the demographic history of species from patterns of polymorphism and divergence observed within and among natural populations. The rationale beyond this approach is that past events, such as a strong bottleneck within a lineage (Gattepaille et al. 2013) or a secondary contact between divergent populations (Durand et al. 2011), leave distinct genetic signatures in gene pools, which can be used to infer the nature and timing of the processes that gave rise to them. For example, after the occurrence of a severe bottleneck, linkage disequilibrium (LD) extends to larger genomic regions than one would expect to find in stable populations (Reich et al. 2001), coalescent trees became star-shaped, and there is a resulting excess of rare mutations across the whole genome. Such patterns of genetic variation have allowed the inference of bottlenecks and subsequent population expansion in recent history of a number of species, including *Drosophila melanogaster* (Li and Stephan 2006), *Arabidopsis thaliana* (François et al. 2008) and *Oryza sativa* (Caicedo et al. 2007).

The rise of approximate Bayesian computation (ABC) (Tavaré et al. 1997; Pritchard et al. 1999; Beaumont et al. 2002) has made a substantial impact on our ability to infer the demographic and evolutionary history of populations from genetic data. ABC is a tool for statistical comparisons between alternative models of evolution (Beaumont et al. 2010). It avoids the difficulty of computing the likelihood function of complex demographic models by comparing the values of a vector of summary statistics between an observed genetic dataset data produced by simulations. By using a rejection-and-regression algorithm

(Beaumont et al. 2002; Blum and François 2009), the posterior probability of a model M can then be approximated by its relative acceptance frequency.

ABC has successfully been used to infer the demographic history and history of evolutionary divergence for a wide range of organisms. For instance, comparing different scenarios of human evolution using ABC, Fagundes et al. (2007) provided unambiguous support for a recent human colonization out of Africa about 50,000 years ago, with no long-term gene flow between continents, thus contributing to the rejection of the 'multiregional hypothesis' (Cavalli-Sforza and Feldman 2003; Lahr 1994; Mellars and Stringer 1989). To understand the history of divergence in the plant genus *Arabidopsis*, Roux et al. (2011) used ABC to reveal a tight relationship between the establishment of a major adaptive mutation in *Arabidopsis* and speciation. Inferring realistic demographic models from genetic data leads to better knowledges about the history of species, but it is also crucial for building statistical null models for the detection of molecular signatures of direct and indirect selection across genomes (Nielsen 2005; Roux et al. 2012).

Coalescent-based ABC has hitherto been mainly applied to haploid and diploid organisms, with little attention given to polyploid species, despite their recognized importance (Jakobsson et al. 2006; St Onge et al. 2012; Lepais et al. 2013). Indeed, polyploidy has played a fundamental role in the evolution and diversification of a wide range of taxa, including invertebrate and vertebrate animals and, especially, plants (Otto and Whitton 2000; Gregory and Mable 2005; Gallardo et al. 2004; Neiman et al. 2011). Concerning plants, the estimated proportion of current species with more than two copies of each chromosome is claimed to reach values as high as 35% (Arrigo and Barker 2012), with its significance having been revealed at very different timescales. At one extreme, for example, comparative genome analysis has revealed a paleopolyploidization event that predates the radiation of angiosperms (Bowers et al. 2003); at the other extreme, several new polyploid species have been reported to have become established in the past 150 years (Soltis et al. 2003). Within a lineage, polyploidization has been proposed as a mechanism of species differentiation by favoring instant sympatric speciation because of the low fitness or sterility of hybrids between diploid parents and their polyploid descendants (Coyne et al. 2004; Wood et al. 2009; Linder and Rieseberg 2004; Mallet 2007). Consistent with this mechanistic, Mayrose et al. (2011) compared speciation

rates between diploids and neo-polyploids in angiosperms and seed-free vascular plants and confirmed that whole-genome duplication (WGD) of diploids has been a major process driving speciation, albeit counterbalanced to some extent by a greater extinction rate of neopolyploids over diploids (Mayrose et al. 2011).

Despite the importance of polyploidization in speciation, significant theoretical and empirical approaches to fit polyploid genomic data to evolutionary or demographic models have emerged only recently (Jones et al. 2013). A pivotal issue concerns determining the effects of genome duplication on the coalescent model (Hudson 1983; Tajima 1983) in order to delimit what we can do with current tools (Hudson 2002; Laval and Excoffier 2004). In this context, a key question concerns how polyploidization will influence demographic inference within a lineage that is likely to have gone through one or more genetic bottlenecks, that may have evolved more than once, and for which (in the case of allopolyploidy) the ancestral population was subdivided into two potentially reproductively isolated units. Not only do these different scenarios potentially complicate demographic inference; they are also intrinsically interesting in themselves. How can recent advances in demographic and evolutionary inference help us to infer the nature and timing of these different events?

One is often interested in distinguishing between the establishment of a polyploid lineage following the simple doubling of the genome within a single population (autopolyploidy) and its establishment as a result of hybridization and genome doubling between two divergent populations or species (allopolyploidy). To model autotetraploid populations, Arnold et al. (2012) recently developed a coalescent framework that assumes tetrasomic inheritance (so that all of the six possible gametes have the same probability of being produced by tetraploid adults) and showed that a rescaled Kingman model for haploids serves as a suitable approximation. Lineages produced by allopolyploidization of course have a more complex history, not least because they combine two genomes that can be potentially quite divergent, each originating from a different species with its own demographic history. The demographic history of an allopolyploid genome thus includes three different histories: the two histories of the contributing parental species, and the history following hybridization and effective genome duplication. Demographic inference involving allopolyploid lineages thus needs to account for, and estimate, the histories of each of these component populations.

There have been substantial recent advances in our understanding of the rôles played by auto- versus allopolyploidisation in speciation between diploid and polyploid species (Jakobsson et al. 2006; Slotte et al. 2008; Wang et al. 2010; St Onge et al. 2012). For instance, St. Onge et al. (2012) used the MIMAR software (Becquet and Przeworski 2007) to distinguish between auto- and allopolyploidization in the divergence between diploid plant *Capsella grandiflora* and its tetraploid derivative *C. bursa-pastoris*, estimating the parameters of the Isolation with Migration (IM) model. In their analysis, they considered both sub-genomes A and B of *C. bursa-pastoris* as effectively independent species experiencing divergent evolution, and compared them with the genome of *C. grandiflora*. They then estimated speciation times between *C. grandiflora* and the *C. bursa-pastoris* A and B genomes, respectively. Although this approach should be straightforward, it is not compatible with modern datasets because it requires the specific assignment of each of the four phased-haplotypes to one of the two diploid sub-genomes (homeologous pairs) for the tetraploid species (but see (Salmon et al. 2012)). MIMAR was originally designed to estimate parameters of the IM model only (the effective population sizes of the ancestral population N_A and for both daughter populations N_1 and N_2 , the time of speciation T_{split} , and migration rates in both directions), and it does not provide an explicit statistical test of alternative scenarios of speciation with polyploidization.

Although the distinction between auto- and allopolyploidy is conceptually useful, it is important to note that notably the distinction between tetrasomic and disomic inheritance between the two modes of polyploidisation are expected to be transitory. Typically, allopolyploid lineages will show disomic inheritance, with preferential pairings between the two homologous chromosomes of each contributing species, i.e., between the homeologues; in this case, the two sub-genomes of an allotetraploid lineage can be seen as two completely isolated diploid populations, because recombination occurs only between the chromosomes contributed by each parent species, and not between them. In contrast, genome duplication leading to autotetraploidy will typically give rise to a lineage with initially tetrasomic inheritance, in which homologous chromosomes will pair at random, and recombination can occur between any of the four chromosome. But tetrasomic inheritance in a recently duplicated tetraploid lineage will gradually give way to disomic inheritance as the polyploid genome becomes progressively diploidised as a result of the divergence between two different homeologous chromosome pairs. Because this process of diploidisation can be

prevented at islands of recombination between the two sub-genomes (Meirmans and Van Tienderen 2013), differences in gene exchange between homeologous pairs across the genome can lead to heterosomy, with some loci showing disomic inheritance and others showing tetrasomic inheritance. Heterosomy has somewhat neglected in the literature (Allendorf and Danzmann 1997; Koning-Boucoiran et al. 2012; Stift et al. 2008; Justin and W. 2002), but it should have profound consequences on the distribution of coalescent times across a genome and thus on demographic and evolutionary inference.

In addition to the mode of origin of a polyploid lineage, an important question concerns the relative timing genome duplication and divergence from its diploid progenitor(s): e.g., does speciation precede genome duplication, or do the two processes coincide (Fig. 1)? A striking observation in plants (Obbard et al. 2006) and animals (Neiman et al. 2011) is that the transition from diploidy to tetraploidy is frequently associated with a shift from obligate outcrossing to predominant self-fertilization (Astaurov 1969; Otto and Whitton 2000). One hypothesis is that this pattern results from the breakdown of self-incompatibility with polyploidization, which the subsequent divergence of the new tetraploid lineage. Alternatively, the subdivision of an ancestral diploid species through colonization may lead to the establishment of self-compatible individuals, e.g., through selection for reproductive assurance (Baker 1955), followed by genome duplication; in this case, genome duplication might occur well after the establishment of self-compatibility because of drift, or because of possible selective advantages conferred by tetraploidy (Chao et al. 2013). To what extent can genetic data allow us to distinguish between these two alternatives using an ABC framework?

In this article, we evaluate the extent to which ABC modelling allows us to recapture the evolutionary and demographic history of polyploid lineages using data (e.g., generated by next-generation sequencing; NGS) with which it is not possible to know either the haplotypic phase of sequence data or the sub-genome from which it originated. Specifically, we use ABC in an attempt to recapture several of the key processes that might have led to the establishment of a tetraploid lineage from one or more progenitor diploid populations, comparing nine different models (Fig. 1). These models vary in: the mode of origin of polyploidy (polyploid speciation, auto- and allopolyploidy); the relative timing of genome duplication and population divergence; and the modes of inheritance characterizing the tetraploid lineage (disomy,

heterosomy or tetrasomy). In addition, we ask both how inferential power might be improved when sequence data from an outgroup are available, and how inference on the basis of a NGS-like dataset (with a large number of loci over few individuals) compares with that based on Sanger sequencing (with fewer loci and more individuals sampled). Finally, we apply our approach to the analysis of polymorphism and divergence from published DNA sequences from diploid and tetraploid lineages of the plant genus *Capsella*, for which hypotheses regarding polyploidization and demographic history have been considered using different approaches (Slotte et al. 2006, 2008; St Onge et al. 2012).

Results and discussion

Distinguishing between allo- and autopolyploidisation

Our simulations show that ABC can reliably distinguish between an allo- or autopolyploid origin of a tetraploid organism on the basis of a realistic genomic dataset (Fig. 2). We generated 1,000 pseudo-observed datasets for each of the allo- and autopolyploidization models and obtained the distributions of relative posterior probabilities consistent with the evolutionary and demographic model used to generate the sequence data. Importantly, our ability to correctly identify the origin of polyploidization depended to some extent on the mode of inheritance assumed for the sequenced loci. Thus, the proportion of successfully supported pseudo-observed datasets decreased with the intensity of recombination between the two sub-genomes of the derived tetraploid lineage. For instance, a scenario of autopolyploidization was correctly identified in 99%, 93.5% and 83.6% of the simulated cases for disomic, heterosomic and tetrasomic inheritance, respectively. This loss of information with increasing heterosomy can be attributed to recombination between the otherwise homeologous chromosomes during the early stages after polyploidization and recovery from the corresponding population bottleneck caused by the genome duplication of a single individual.

Distinguishing between an auto- and allopolyploid origin of a polyploid lineage has hitherto required detailed cytogenetic analysis. Such work included the extraction and staining of metaphasic chromosomes in order to characterize chromosomal configurations, such as the telomeric heterochromatin in the long arm (Tymowska and Fischberg 1982; Chalup et al. 2012), and cluster analysis to test whether each of the homeologous chromosomes group together as expected under autopolyploidy (Gutiérrez et al. 1994). Despite improvements in the acquisition of molecular markers, the rationale beyond this test has more or less remained unchanged, with auto- versus allopolyploid origins still distinguished by testing for the occurrence of divergent chromosomes within the same tetraploid cytotype, even where data from parental species are not available (Mráz et al. 2012). The mode of inheritance discerned by cytogenetic analysis is still frequently used to directly infer the polyploid origin (Hardy et al. 2000), even though this relationship may be misleading (Le Comber et al. 2010). Even in recent studies using simulation-based methods, the mode of

inheritance is assumed to be disomic by default, so as to infer the origin of polyploidization by way of estimates of divergence dates between each sub-genome pair and putative diploid relatives (St Onge et al. 2012). Our analysis now points to ABC as a powerful means of inference for the origin of polyploid lineages, not least because it is able to account for the demographic history and the diversity in recombinational models between homeologous chromosomes.

Inferring the mode of inheritance

Our simulations also show that, beyond the fundamental distinction between allo- and autopolyploidisation, it is possible to infer (rather than assume) the mode of inheritance for a polyploid lineages (Fig. 3). Thus, we were able to correctly attribute the mode of inheritance as either disomic, heterosomic or tetrasomic at least 96.9% of the time, regardless of the specific scenario underlying speciation. This result should give us confidence in the statistical support for any model of inheritance provided by ABC on the basis of sequence variation across the genome, without needing to deal with logistically more demanding and complex cytogenetic approaches.

Historically, the approaches to establish the mode of inheritance have been based on intensive cytogenetic studies or segregation analysis from experimental crosses between parents of known genotypes (Kamiri et al. 2011; Pupilli et al. 1997). By using segregation patterns of genetic markers in offspring at multiple loci, it has been necessary to compute the probabilities of the different models of inheritance to finally test the support of the best hypothesis through arbitrary interpretations of the Bayes factor value (Olson 1997). Although recent technological advances have made hundreds of molecular markers available for screening the genomes of non-model species (Gayral et al. 2013), it will often be difficult and/or prohibitively time-consuming to obtain and genotype sufficiently large numbers of progeny for most wild species. Catalàn et al. (2006) have shown that, in some cases, it might be possible to circumvent this problem by making use of microsatellite profiles from natural populations to discriminate between disomic and tetrasomic inheritance without performing crosses, but this approach will rarely be feasible. Some methods have been successfully adopted to accommodate the possibility of intermediate heterosomic models of inheritance, but these, too, make use information from a segregation analysis (Stift et al. 2008) and so will

often not be available.

As our study shows, investigating the mode of inheritance can be made using ABC on the basis of DNA sequences, without having to assume a strict relationship between the historical scenario and any biological process. Although we have considered probably the least complex scenarios that one might face, the approach we have used should easily be extended or adapted to accommodate more complex models. A first extension, for example, would be the implementation of a framework to identify a gradual shift from polysomy to disomy, since multivalent pairing is ephemeral (Otto 2007). Such models would provide a better description of the linear or quadratic relationship between the time of polyploidization and the proportion of the genome that has become disomic, but it would also provide a better understanding of the evolutionary forces driving such transitions.

Testing for the relative timing of duplication and speciation

It is clear from our simulations that ABC can provide a robust test for the co-occurrence of polyploidization and speciation under certain inheritance modes (Fig. 4). We found that ABC is frequently able to recapture the evolutionary and demographic parameters for $T_{\text{split}} = T_{\text{WGD}}$ (polyploid speciation) and $T_{\text{split}} > T_{\text{WGD}}$ for cases with disomic (with respectively 99% and 96.8% of correctly supported pseudo-observed datasets) and heterosomic inheritance (with respectively 96.8% and 92.4% of correctly supported pseudo-observed datasets). However, with recombination between homoeologous chromosomes, power to recapture the simulated parameters fell dramatically (to values of 58.3% and 55.5% for $T_{\text{split}} = T_{\text{WGD}}$ and $T_{\text{split}} > T_{\text{WGD}}$, respectively). Under fully tetrasomic inheritance, distinguishing between these two models is similar to estimating the timing of speciation between two undifferentiated gene pools linked by elevated introgression. It is thus clear that a minimum level of preferential pairing between homoeologous chromosomes in the genome will be required for accurately testing for the co-occurrence of polyploidization and speciation. It will thus be important to test for the mode of inheritance prior to attempting such evaluations.

Of course, polyploidization represents one of the few realistic causes of sympatric speciation, and the coincidence of polyploidisation and speciation might be expected by default for a number of reasons; its statistical rejection is thus likely to be the more interesting outcome of ABC modeling. A newly formed

polyploid population can be isolated from the parental species as a result of induced genomic incompatibilities between populations with different ploidy levels (Fowler and Levin 1984; Rodriguez 1996) and the rapid reorganization of chromosomes that result from “genomic shock” (Parisod et al. 2010). Polyploidization can also induce speciation as a result of a change in the mating system, especially the evolution of self-fertilization (Otto and Whitton 2000; Soltis et al. 2010; Pannell et al. 2004). Finally, genome doubling can extend the range of trait values, allowing the colonization of new ecological areas and leading to a geographical isolation and allopatry (Levin 1983). Cases where polyploidisation does not co-occur with speciation are likely to be rare, but testing for them using ABC modelling is straightforward.

Inferring model parameters

We also considered the precision of ABC to retrieve parameter values for simulated models when $T_{\text{split}} = T_{\text{WGD}}$. Figure 5 shows the distributions of the ratio $\beta = \theta_{\text{i-est}} / \theta_{\text{i-real}}$, where $\theta_{\text{i-est}}$ is the median of the estimated posterior distribution of the parameter θ_i and $\theta_{\text{i-real}}$ is the true parameter value used to produce the pseudo-observed dataset. For the three different modes of inheritance, the distributions of β were generally centered around one, pointing to high accuracy in parameter estimation for all parameters and models. However, there were some differences in precision among the parameters and models measured by the median absolute deviation to the median of β (MAD_β). Concerning the time of speciation (T_{split}), the effective population sizes of the diploid (N_1) and ancestral species (N_a), and the proportion of genome segregating with disomic inheritance under the heterosomic model (α), ABC provides very precise and trustworthy estimates, with small MAD_β (Fig. 5). In contrast, the more elevated MAD_β measured for the effective population size of the tetraploid species (N_2) suggests that we should have less confidence in this parameter estimate: caution is thus warranted when interpreting posterior distributions of N_2 for all modes of inheritance when $T_{\text{split}} = T_{\text{WGD}}$ (Fig. S1-S3).

Our results suggest that ABC should provide a useful tool for the joint estimation of a number of key parameters in the process of polyploidisation, such as T_{split} , T_{WGD} and α , but estimates of N_2 should be interpreted with caution, and estimates of the effective population size of the ancestral diploid population prior to tetraploidization (N_{2a}) have little value. When $T_{\text{split}} > T_{\text{WGD}}$ (Fig. 6), ABC provided precise estimates

for N_1 , N_a , α and T_{split} . The precision achieved when estimating T_{WGD} and N_2 were globally credible for all models, with a few caveats concerning their estimates for disomic and heterosomic autopolyploids. As might be expected, the strong bottleneck occurring at the time of polyploidization affects estimates of the effective population size of the diploid lineage before the genome duplication (N_{2a}), as most of the informative polymorphism is lost as a result of the bottleneck. Thus, the precision achieved for estimates of N_{2a} was between about 8.5 (disomic autopolyploid) and 18 (heterosomic allopolyploid) times worse than that achieved for N_1 , as measured by MAD_β . This lack of statistical support is evident more in terms of very wide posterior distributions than by well-defined unimodal posteriors around wrong values, i.e., it is a question of reduced precision rather than accuracy (Fig. S4-S9).

Genome duplication has important consequences for genome structure, gene content and gene expression (Doyle et al. 2008). Inferring its date for a given biological model is therefore a key feature contributing to our understanding of molecular evolution in polyploids. In the simplest case, when polyploidization co-occurred with speciation, this date can ultimately be estimated using a molecular clock on the basis of the silent-site divergence between the polyploid species and its closely related putative diploid progenitor (Doyle and Egan 2010; Senchina et al. 2003). However, even for this simple model, the inferred date more reflects the coalescent time between diploid and polyploid lineages than the time of population subdivision (Wakeley 2009). Thus, for large ancestral population sizes, variation in coalescence depth across the genome can be large enough to complicate the use of molecular clock approaches, whereas ABC precisely provides estimate for the population separation when $T_{\text{split}} = T_{\text{WGD}}$. If genome duplication has occurred after speciation (looking forward in time), the use of molecular divergence between duplicated genes within the tetraploid (Lynch and Conery 2000) appears to be a reasonable approach to infer the time of polyploidization, notably for allopolyploidisation. However, in addition to its weakness in dealing with tetrasomic inheritance in autopolyploids, this approach will also be highly sensitive to incomplete lineage sorting when the difference between the time of speciation and the time of polyploidization is small compared with the ancestral N_{2a} .

The relative efficiency of contrasting sequencing methods

We also considered how best to optimize genome sampling for model choice and parameter estimation. Overall, our results indicate that, for demographic and evolutionary inference, the number of surveyed loci is more important than having a sequenced outgroup or a larger number of sequenced individuals. With an NGS-like dataset (e.g., a large number of short contigs from an Illumina run), the absence of information from a putative outgroup had only a very small negative effect on our ability to distinguish between auto- and allopolyploidisation (Table 1), to identify correctly the mode of inheritance (Table 2), or to test for co-occurrence (or not) of T_{split} and T_{WGD} (Table 3). Precision in parameter estimates were also highly similar with or without a sequenced outgroup, either for the three $T_{\text{split}} = T_{\text{WGD}}$ scenarios (Table 4), or for the six auto- and allopolyploid scenarios, with very little gained in terms of MAD_{β} by including an outgroup (Table 5). It would seem from our results that the effort in acquiring data from an outgroup species is unlikely to be worthwhile if the main objective is inferring the demographic history of a diploid and derived tetraploid species – though other benefits of including an outgroup will need to be evaluated, such as determining the polarity of mutations, the quantification of fixed alleles in both lineages instead of relying on the total number of fixed differences between them (Ramos-Onsins et al. 2004), the estimation of locus-specific mutation rates.

With data that might be obtained via Sanger sequencing, we found in most cases our inferences remained robust to decreases in the number of surveyed loci. In particular, testing for auto- versus allopolyploidization was still reliable for a scenario of disomic inheritance, because the fixed retention of both parental genome copies represents an unambiguous signature detected by ABC (Table 1). The robustness of this model comparison decreased for increased levels of tetrasomy in genomes, regardless of the presence of an outgroup sequence (Table 1). ABC correctly distinguished between disomic and tetrasomic models of inheritance for all of the three scenarios of speciation (Table 2); when the wrong model was inferred, it tended to be in favor of heterosomy.

Sanger-like datasets were sufficient to test for the co-occurrence of speciation and polyploidization for tetraploids with disomic inheritance (Table 3), but a greater number of loci was required for species with heterosomic inheritance. As for NGS-like datasets, this test is also not conceivable when the biological model has a polysomic inheritance. Whereas reducing the number of sampled loci had only a small negative

impact on model-comparisons, such datasets were not reliable for estimating key parameters in the demographic and evolutionary process. ABC is thus only poorly able to estimate T_{split} and T_{WGD} from Sanger-like datasets for scenarios where $T_{\text{split}} = T_{\text{WGD}}$ (Table 4), or to distinguish between auto- and allopolyploidy (Table 5). In many study systems, of course, no suitable outgroup will be available, and Sanger-like datasets that have already been generated are sometimes the only sources of information. ABC can profitably be performed on these data to infer polyploid history, particularly for model comparisons, but it will be weak for parameter estimation.

Empirical application

As a brief case study, we evaluated the use of our approach for inferring the evolutionary and demographic history of divergence between diploid *Capsella rubella* and tetraploid *C. bursa-pastoris*, for which orthologous sequences have been published (Slotte et al. 2008; St Onge et al. 2012). First we evaluated the mode of inheritance of homeologous chromosomes in tetraploid *C. bursa-pastoris* and found that, for each of the scenarios involving simultaneous polyploidy and speciation, allopolyploidization and autopolyploidization there was always strong support for a model of disomic inheritance (Table 6-a). This result confirms conclusions based on previous analysis (Slotte et al. 2006; Hurka et al. 1989; Hurka and Neuffer 1997).

Second, we statistically evaluated the relative posterior probabilities of the three alternative demographic scenarios under disomic, heterosomic and tetrasomic inheritance. Here, we found that, in all three cases, a model of allopolyploidization best explains the origin of *C. bursa-pastoris* (Table 6-b). This unambiguous support of a model of allopolyploidy with disomic inheritance was also found when we compared all nine models in the same analysis, with a posterior probability ≈ 0.82 (Fig. 1). The parameters estimated for this model (Fig. 7) point to an ancestral split between the two diploid lineages that ultimately contributed to *C. bursa-pastoris* $\approx 735,371$ years ago (with a higher posterior density at 95, 95HPD, estimated between 266,414 and 1,521,383 years), and that they then hybridized to form *C. bursa-pastoris* $\approx 89,199$ years ago (95HPD between 34,064 and 198,738 years).

Assuming a model with constant population size since the ancestral split, the lineage leading to *C.*

rubella had an effective size of $\approx 85,000$ individuals (with 95HPD between 24,966 and 205,524 individuals). The estimated current population size of each sub-genomes of *C. bursa-pastoris* ($\approx 355,900$ individuals; 95HPD between 138,241 and 863,222) could not be compared directly to the long-term effective population size of *C. rubella*, because we infer that population size grew rapidly at a rate of 1.343×10^{-4} . Thus, the long-term effective population size for each of the *C. bursa-pastoris* sub-genomes is better approximated by its geometric mean rather than by the current estimate, providing a value of ≈ 842 individuals over the last $\approx 90,000$ years. Given that the effective population size of a diploid lineage immediately prior to whole-genome duplication cannot be faithfully estimated (Table 5, and see above), we do not provide the posterior distribution of this parameter. However, overall, a goodness-of-fit test indicates that our inferences provide a satisfactory reflection of the empirical dataset, yielding the expected distribution of summary-statistics simulated from the joint posterior distribution of the parameters under the best-supported model (Fig S10; Cornuet et al. 2010).

Conclusions

Our study establishes the value of ABC for inferring the evolutionary and demographic history of polyploid species on the basis of genomic data generated by state-of-the-art sequencing technologies. Not only does our approach allow us to recapture with satisfying accuracy some of the parameters used to simulate genomic data during the divergence and coalescence of lineages in the evolution of polyploid complexes; we also find that it can be applied to real data, making plausible inferences. It is particularly satisfying to note that NGS data can be used for the analyses we have tested, without any need to attribute alleles to the potentially divergent homoeologous genomes making up a polyploid lineage. Unsurprisingly, our ability to infer certain parameters was weakened for scenarios involving tetrasomic inheritance; such inheritance is analogous to effectively free gene flow between spatially isolated populations, which we should not expect to differ: tetrasomic inheritance erodes signatures of genomic subdivision and hence our ability to infer the origin of the constitutive genomes just as migration erodes our ability to attribute alleles to one deme or another. Nevertheless, it is important to note that a certain degree of disomic inheritance in a genome that displays heterosomy quickly becomes sufficient to allow accurate inference. It is also worth

emphasizing that a great deal can be achieved by increasing the numbers of loci sampled, particularly for genomes displaying heterosomy. Of course, actual scenarios of polyploid are often a great deal more complex than those we have explored here, not least in complexes with higher ploidy levels and mixtures of auto- and allopolyploidy. It should, however, be straightforward to extend our approach for dealing with other more complex scenarios. Indeed, the flexibility of ABC allows the testing of quite specific models of evolution and demography. How additional model complexity in the context of polyploid evolution will affect our confidence in parameter estimation awaits further work.

METHODS

Models

We assessed the ability of ABC as a robust approach to investigate polyploidization by surveying a set of nine evolutionary and demographic scenarios (Fig. 1). These scenarios differ in terms of their origin of genome doubling (autopolyploid or allopolyploid), the timing of polyploidization (whole-genome duplication, WGD, is either coincident with speciation for models I, or occurred after speciation for models II), and the mode of inheritance of chromosomes (disomic, heterosomic, tetrasomic). All of the nine models describe the subdivision of an ancestral diploid population at time T_{split} into two populations: the first population remains diploid with a constant population size N_1 , while the second population experiences WGD at the time T_{WGD} . We treated the sub-genomes of the polyploid species as two discrete populations, 2A and 2B, each sharing the same current effective population size N_2 . Since we assume a strong bottleneck in population size due to drastic founding effect of the new tetraploid lineage, both 2A and 2B also share the same exponential growth rate computed as $-\log(1/N_2)/T_{\text{WGD}}$.

To simulate disomic inheritance, we considered a scenario in which there was no gene flow between 2A and 2B from time of genome duplication. As for every model, both 2A and 2B sub-genomes are also assumed to be isolated from the diploid species. We approximated the tetrasomic inheritance following Meirmans and Van Tienderen (2013) by attributing a rate of effective migration arbitrarily fixed to 10 individuals per generation, which is sufficiently high to prevent differentiation between sub-genomes (Whitlock and McCauley 1999; Wang 2004; Templeton 2006). It is worth noting that this approximation only holds up because we assumed a single panmictic tetraploid population that receives no migrants from unsampled populations. Heterosomy is implemented in our simulations as a binomial distribution of size equal to the number of investigated loci, and a probability for a given locus to exchange alleles between sub-genomes equal to α . Thus, for heterosomic models, α represents the expected proportion of the genome experiencing tetrasomic inheritance, and $(1-\alpha)$ represents the expected proportion of the genome experiencing disomic inheritance.

In allopolyploid models, the diploid species and the 2-A lineage first come together at time T_{WGD} , following which both of the ancestral diploid lineage and the 2-B lineage join at time T_{split} . In autopolyploid

models, the first fusion backward in time occurred between 2-A and 2-B at time T_{WGD} , then the ancestral tetraploid lineage with effective population size $N_{2\text{-anc}}$ is joined to the first lineage at time T_{split} .

Finally, $T_{\text{split}}=T_{\text{WGD}}$ for the three models I, and $T_{\text{split}}>T_{\text{WGD}}$ for the six models II.

Coalescent simulations

For each model, we obtained a set of 5×10^6 multilocus simulated data sets using *msnsam* (Ross-Ibarra et al. 2008), a modified version of the *ms* software (Hudson 2002), allowing for variation in sample size among loci under an infinite-site mutation model. The command lines used to simulate the different models are shown in supplemental text S1. We used the software *PriorGen* to generate the prior distributions given the nine models, and *mscalc* to compute the summary statistics described below from the output of *msnsam* (Ross-Ibarra et al. 2008, 2009; Roux et al. 2011, 2013).

We used a wide uniform prior distribution [0-10] common to all models and shared by all parameters in order to consistently explore a large panel of realistic biological scenarios. Values for α have been randomly sampled in the uniform prior [0.025-0.975]. To convert coalescent units in natural demographic units, effective population sizes N_A , N_1 , $N_{2\text{-anc}}$ and $N_{2\text{-A}}$ (or $N_{2\text{-B}}$) have to be multiplied by the effective population size of the virtual reference population arbitrarily fixed to 10^5 . For example, following the used prior for N_1 , this parameter can take a value between zero and one million. T_{split} and T_{WGD} have to be multiplied by $4 \cdot 10^5$ to obtain the number of generations in demographic units, so that T_{split} was sampled between zero and four million generations.

For each simulated dataset, we computed an array of summary statistics related to polymorphism and divergence, as is widely used in the literature (Wakeley and Hey 1997; Becquet and Przeworski 2007; Ross-Ibarra et al. 2008, 2009; Roux et al. 2011, 2013; Fagundes et al. 2007): Tajima's π (Tajima 1983), Watterson's θ_W (Watterson 1975), F_{ST} estimated as $1 - \pi_S / \pi_T$ (where π_S is the mean pairwise diversity for both species and π_T is the pairwise diversity calculated from the total alignment), Tajima's D (Tajima 1989), total interspecific divergence Div , and the net interspecific divergence $netDiv$ ($=Div - \pi_S$). We add two alternative sets of summary statistics depending on the availability of a sequenced outgroup for the user. In cases where a sequenced outgroup was assumed to be available to orientate mutations (Ramos-Onsins et al. 2004), we also

used as summary statistics the number of derived biallelic positions exclusively polymorphic in the diploid species (Sx_A) and the number of derived positions exclusively polymorphic in the tetraploid species (Sx_B). Concerning the latter category, a sequenced tetraploid individual is seen as the merging of two copies from sub-genome 2-A and two copies from sub-genome 2-B without the possibility of distinguishing the two. We counted the number of polymorphic sites shared by both species (Ss), *i.e.*, a biallelic polymorphism found in the diploid species as well as in the merged 2-A and 2-B tetraploid species. We computed the number of positions for which a derived allele is fixed in the first species and the ancestral allele is fixed in the second one (Sf_A). In the same way, we defined a Sf_B site as a genomic position for which an ancestral allele is fixed in the diploid species and the derived allele is fixed in both 2-A and 2-B subgenomes. The position Sf_AxB (or Sf_BxA) defines a derived allele fixed in the diploid species (or in the tetraploid), but still segregating in the tetraploid species (or in the diploid). In the situation where a sequenced outgroup was assumed not to be available to the investigator, we computed Sx_A as the total number of polymorphic positions specific to the diploid species regardless of the fixed allele in the tetraploid species, and used the same rational to compute Sx_B . In this situation with no outgroup sequence, Sf is equal to the sum of Sf_A and Sf_B . Finally, we used the mean and the standard deviation of each quantities computed over all of the loci as summary statistics.

Alternative sequencing strategies

We used coalescent simulations to assess the robustness of ABC through four alternative experimental schemes: datasets with or without an outgroup; and data generated with NGS-like or Sanger-like sequencing. While the presence or absence of an outgroup only impacts the summary statistics described above, the sequencing strategy mainly impacts the sampling size as well as the number of re-sequenced loci. For the Sanger-like dataset, we assumed the availability of a set of 30 loci of mean synonymous length equal to 110 (SD=10) resequenced for around 60 sampled chromosomes (SD=10); this is similar to Sanger-like datasets recently analysed in the literature (Roux et al. 2011). For the NGS-like dataset, we assumed a set of 852 loci of mean synonymous length equal to 70 (SD=30) resequenced for around 16 sampled chromosomes (SD=4); this is similar to the dataset investigated in Roux et al. (2013). For each of the four experimental strategies, we simulated 1,000 pseudo observed datasets under the nine scenarios investigated.

ABC analysis

Model selection

For each model comparison, we evaluated the relative posterior probabilities of the alternative models using a feed-forward neural network, implementing a nonlinear multivariate regression by considering the model itself as an additional parameter to be inferred with the R package “abc” (Csillery et al. 2012). The 0.025% replicate simulations nearest to the observed values for the summary statistics were selected, and these were weighted by an Epanechnikov kernel that reaches a maximum when $S_{\text{obs}}=S_{\text{sim}}$, where S_{obs} are the summary statistics computed from the observed or the pseudo observed datasets. Computations were performed using 50 trained neural networks and 10 hidden networks in the regression.

Parameter estimation

Posterior distributions of the parameters describing the best model were estimated using a nonlinear regression procedure. Parameters were first transformed according to a log-tangent transformation (Hamilton et al. 2005). Only the 2,000 replicate simulations that provided the smallest associated Euclidean distance $\delta = \|S_{\text{obs}} - S_{\text{sim}}\|$ were considered. The joint posterior parameter distribution was then obtained by means of weighted nonlinear multivariate regressions of the parameters on the summary statistics. For each regression, 50 feed-forward neural networks and 20 hidden networks were trained using the R package “abc”.

Capsella dataset

We downloaded published sequences obtained in *C. rubella* and *C. bursa-pastoris* (St Onge et al. 2012; Slotte et al. 2008) from GenBank (accession numbers EF683687– EF684898 and JQ418636–JQ419488), making a total of 13 aligned loci. Only bi-allelic polymorphism at silent sites (introns + synonymous positions from exons) were considered in the analysis, based on the GenBank annotations. For each of the nine alternative models, we ran three million multilocus simulations, with the following uniform prior distributions. $N_{C. rubella}$ [0-1,000,000]; $N_{C. bursa-pastoris}$ [0-1,000,000]; N_{ancestor} [0-2,000,000]; T_{split} [0-4,000,000]; T_{WGD} [0- T_{split}]. A mutation rate of 1.5×10^{-8} per site and per generation was then used to rescale the coalescent

units in demographic units, assuming a generation time of one year. We finally applied the same ABC procedure as described above. Simulations and ABC analysis for the *Capsella* dataset were run for a total of three hours when parallelized on a cluster using 150 CPUs.

DATA ACCESS

The used code is available at: http://www.unil.ch/webdav/site/dee/shared/softs/ABC_WGD.tar.gz

ACKNOWLEDGMENTS

We are grateful to Jonathan Wendel for suggestions and comments the manuscript. The computations were performed at the Vital-IT Center for high-performance computing of the SIB Swiss Institute of Bioinformatics (<http://www.vital-it.ch>). CM was supported by a postdoctoral fellowship of the Swiss National Science Foundation awarded to JRP.

FIGURE LEGENDS

Figure 1. Alternative scenarios of speciation and polyploidization.

Parameters are as follows. N_1 : effective population size of the current diploid species. N_2 : effective population size of each sub-genomes making the tetraploid species. N_{2A} : effective size of the diploid population ancestral to the tetraploid species, for autopolyploidization and allopolyploidization scenarios only. N_A : effective population size of the common ancestor of diploid and tetraploid species. T_{split} : time of speciation. T_{WGD} : time of whole-genome duplication. α : proportion of the genome associated to tetrasomic inheritance, for heterosomic species only.

Numbers are the relative posterior probabilities attributed by ABC from the *Capsella* dataset when all models were compared in the same analysis.

The red rectangle indicates the best-fitting model to the *Capsella* dataset.

Figure 2: Empirical distributions of relative posterior probabilities to correctly support the polyploid origin regarding the mode of inheritance.

Red areas represent the distributions of the estimated values of $P(\text{Auto} | \text{Auto})$ and $P(\text{Allo} | \text{Allo})$ obtained after model-comparisons using ABC from 1,000 pseudo-observed datasets randomly simulated under auto- and allopolyploidization scenarios respectively. Pairwise model-comparisons were made for disomic, heterosomic and tetrasomic modes of inheritance. The black boxes within the red areas delimit the 25% and 75% quantiles of each distribution, and the white dots show the median. In pairwise model-comparisons, a model M is the best-supported model if its relative posterior probability is greater than 0.5.

Figure 3: Empirical distributions of relative posterior probabilities to correctly support the mode of inheritance regarding the polyploid origin.

Red areas represent the distributions of the estimated values of $P(\text{Disomy} | \text{Disomy})$, $P(\text{Heterosomy} | \text{Heterosomy})$ and $P(\text{Tetrasomy} | \text{Tetrasomy})$ obtained after model-comparisons using ABC from 1,000 pseudo-observed datasets randomly simulated under disomic, heterosomic and tetrasomic modes of inheritance, respectively. Model-comparisons were made for auto- and allopolyploidization scenarios,

respectively.

Figure 4: Empirical distributions of relative posterior probabilities to correctly support the co-occurrence of speciation and polyploidization regarding the mode of inheritance.

Red areas represent the distributions of the estimated values of $P(T_{\text{split}}=T_{\text{WGD}} \mid T_{\text{split}}=T_{\text{WGD}})$ and $P(T_{\text{split}}>T_{\text{WGD}} \mid T_{\text{split}}>T_{\text{WGD}})$ obtained after model-comparisons using ABC from 1,000 pseudo-observed datasets randomly simulated when $T_{\text{split}}=T_{\text{WGD}}$ and $T_{\text{split}}>T_{\text{WGD}}$ respectively. Pairwise model-comparisons were made for disomic, heterosomic and tetrasomic mode of inheritance. The black boxes within the red areas delimit the 25% and 75% quantiles of each distribution, and the white dot show the median.

Figure 5: Estimates provided by ABC from pseudo-observed datasets simulated when $T_{\text{split}}=T_{\text{WGD}}$, regarding the mode of inheritance.

The y-axis shows the median of the estimated posterior distribution divided by the true value for each parameter. The results are for 500 pseudo-observed datasets simulated under the disomic (Fig. S1), heterosomic (Fig. S2) and tetrasomic (Fig. S3) modes of inheritance when $T_{\text{split}}=T_{\text{WGD}}$.

Figure 6: Estimates provided by ABC from pseudo-observed datasets simulated for auto- and allopolyploidization, regarding the mode of inheritance.

The y-axis shows the median of the estimated posterior distribution divided by the true value for each parameter. The results are for 500 pseudo-observed datasets simulated under the disomic (Fig. S4 and Fig S7), heterosomic (Fig. S5 and Fig S8) and tetrasomic (Fig. S6 and Fig S9) modes of inheritance, and for auto- and allopolyploidization respectively.

Figure 7: Parameter estimates of the best supported model of disomic-allopolyploidization.

Prior and posterior distributions for parameters are open and shaded respectively.

The median of each parameter is given along, converted into number of effective individuals or years.

Figure S1: Comparison between posterior distribution and real value for the disomic inheritance when $T_{\text{split}}=T_{\text{WGD}}$ for 500 pseudo-observed datasets.

For each parameter, the red vertical line indicates the real parameter value used to simulate the pseudo-observed dataset, and the histogram shows the estimated posterior distribution obtained using ABC.

Figure S2: Comparison between posterior distribution and real value for the heterosomic inheritance when $T_{\text{split}}=T_{\text{WGD}}$ for 500 pseudo-observed datasets.

For each parameter, the red vertical line indicates the real parameter value used to simulate the pseudo-observed dataset, and the histogram shows the estimated posterior distribution obtained using ABC.

Figure S3: Comparison between posterior distribution and real value for the tetrasomic inheritance when $T_{\text{split}}=T_{\text{WGD}}$ for 500 pseudo-observed datasets.

For each parameter, the red vertical line indicates the real parameter value used to simulate the pseudo-observed dataset, and the histogram shows the estimated posterior distribution obtained using ABC.

Figure S4: Comparison between posterior distribution and real value in cases of autopolyploidization with a disomic inheritance, over 500 pseudo-observed datasets.

For each parameter, the red vertical line indicates the real parameter value used to simulate the pseudo-observed dataset, and the histogram shows the estimated posterior distribution obtained using ABC.

Figure S5: Comparison between posterior distribution and real value in cases of autopolyploidization with a heterosomic inheritance, over 500 pseudo-observed datasets.

For each parameter, the red vertical line indicates the real parameter value used to simulate the pseudo-observed dataset, and the histogram shows the estimated posterior distribution obtained using ABC.

Figure S6: Comparison between posterior distribution and real value in cases of autopolyploidization with a tetrasomic inheritance, over 500 pseudo-observed datasets.

For each parameter, the red vertical line indicates the real parameter value used to simulate the pseudo-observed dataset, and the histogram shows the estimated posterior distribution obtained using ABC.

Figure S7: Comparison between posterior distribution and real value in cases of allopolyploidization with a disomic inheritance, over 500 pseudo-observed datasets.

For each parameter, the red vertical line indicates the real parameter value used to simulate the pseudo-observed dataset, and the histogram shows the estimated posterior distribution obtained using ABC.

Figure S8: Comparison between posterior distribution and real value in cases of allopolyploidization with a heterosomic inheritance, over 500 pseudo-observed datasets.

For each parameter, the red vertical line indicates the real parameter value used to simulate the pseudo-observed dataset, and the histogram shows the estimated posterior distribution obtained using ABC.

Figure S9: Comparison between posterior distribution and real value in cases of allopolyploidization with a tetrasomic inheritance, over 500 pseudo-observed datasets.

For each parameter, the red vertical line indicates the real parameter value used to simulate the pseudo-observed dataset, and the histogram shows the estimated posterior distribution obtained using ABC.

Figure S10: Principal component analysis of summary statistics simulated from the posterior distribution under the best-supported model.

A principal component analysis is performed in the space of summary statistics after 10,000 simulated data sets using the posterior distribution under the allopolyploid scenario with disomic inheritance (blue dots). The observed *Capsella* target is added on each plane of the PCA (red dot).

TABLES**Table 1: Robustness of ABC for distinguishing between auto- and allopolyploidization.**

Numbers represent the proportions of pseudo-observed datasets supported by ABC as having an auto- or allopolyploid origin over pairwise model-comparisons. For each combination of {mode of inheritance} x {scenario} x {sequencing strategy} x {outgroup availability}, 1,000 pairwise model-comparisons have been made.

Table 2: Robustness of ABC for estimating the mode of inheritance.

Numbers represent the proportions of pseudo-observed datasets supported as having a disomic, heterosomic or tetrasomic inheritance over model-comparisons by ABC. For each combination of {scenario} x {sequencing strategy} x {outgroup availability} x {mode of inheritance}, 1,000 pairwise model-comparisons were conducted.

Table 3: Robustness of ABC for testing for the co-occurrence of speciation and polyploidization.

Numbers represent the proportions of pseudo-observed datasets supported by ABC as $T_{\text{split}}=T_{\text{WGD}}$ and $T_{\text{split}}>T_{\text{WGD}}$ over model-comparisons. For each combination of {mode of inheritance} x {scenario} x {sequencing strategy} x {outgroup}, 1,000 pairwise model-comparisons were conducted.

Table 4: Precision in parameter estimates under different conditions when $T_{\text{split}}=T_{\text{WGD}}$.

Reported values are the median value of the distribution of β (ratio between the estimated and the true values for each parameter producing the 500 independently random pseudo-observed datasets), and its median absolute deviation to the median (MAD).

Table 5: Precision in parameter estimates under different conditions for cases of auto- and

allopolyploidization.

Reported values are the median value of the distribution of β (ratio between the estimated and the true values for each parameter producing the 500 independently random pseudo-observed datasets), and its median absolute deviation to the median (MAD).

Table 6: Relative posterior probabilities computed along the three different analyses for the *Capsella* dataset.

- a) Relative posterior probabilities of *C. bursa-pastoris* being disomic, heterosomic and tetrasomic, within different demographic scenarios (polyploid speciation, allopolyploidization, autopolyploidization)
- b) Relative posterior probabilities of *C. bursa-pastoris* being associated to different demographic scenarios within different recombinational regimes.

REFERENCES

- Allendorf FW, Danzmann RG. 1997. Secondary Tetrasomic Segregation of MDH-B and Preferential Pairing of Homeologues in Rainbow Trout. *Genetics* **145**: 1083–1092. <http://www.genetics.org/content/145/4/1083.short> (Accessed November 11, 2013).
- Arnold B, Bomblies K, Wakeley J. 2012. Extending coalescent theory to autotetraploids. *Genetics* **192**: 195–204. <http://www.genetics.org/content/192/1/195.short> (Accessed August 7, 2013).
- Arrigo N, Barker MS. 2012. Rarely successful polyploids and their legacy in plant genomes. *Curr Opin Plant Biol* **15**: 140–6. <http://dx.doi.org/10.1016/j.pbi.2012.03.010> (Accessed August 20, 2013).
- Astaurov BL. 1969. Experimental Polyploidy in Animals. *Annu Rev Genet* **3**: 99–126. <http://www.annualreviews.org/doi/abs/10.1146/annurev.ge.03.120169.000531> (Accessed September 27, 2013).
- Baker HG. 1955. Self-compatibility and establishment after 'long-distance' dispersal. *Evolution (N Y)* **9**: 347–349.
- Beaumont M a., Nielsen R, Robert C, Hey J, Gaggiotti O, Knowles L, Estoup A, Panchal M, Corander J, Hickerson M, et al. 2010. In defence of model-based inference in phylogeography. *Mol Ecol* **19**: 436–446. <http://dx.doi.org/10.1111/j.1365-294X.2009.04515.x>.
- Beaumont MA, Zhang W, Balding DJ. 2002. Approximate Bayesian computation in population genetics. *Genetics* **162**: 2025–2035.
- Becquet C, Przeworski M. 2007. A new approach to estimate parameters of speciation models with application to apes. *Genome Res* **17**: 1505–1519. <http://dx.doi.org/10.1101/gr.6409707>.
- Blum MGB, François O. 2009. Non-linear regression models for Approximate Bayesian Computation. *Stat Comput* **20**: 63–73. <http://link.springer.com/10.1007/s11222-009-9116-0> (Accessed May 23, 2013).
- Bowers JE, Chapman BA, Rong J, Paterson AH. 2003. Unravelling angiosperm genome evolution by phylogenetic analysis of chromosomal duplication events. *Nature* **422**: 433–8. <http://dx.doi.org/10.1038/nature01521> (Accessed September 11, 2013).
- Caicedo AL, Williamson SH, Hernandez RD, Boyko A, Fledel-Alon A, York TL, Polato NR, Olsen KM, Nielsen R, McCouch SR, et al. 2007. Genome-wide patterns of nucleotide polymorphism in domesticated rice. ed. G. McVean. *PLoS Genet* **3**: 1745–56. <http://dx.plos.org/10.1371/journal.pgen.0030163> (Accessed August 23, 2013).
- Cavalli-Sforza LL, Feldman MW. 2003. The application of molecular genetic approaches to the study of human evolution. *Nat Genet* **33**: 266–275.
- Chalup L, Grabiele M, Solís Neffa V, Seijo G. 2012. Structural karyotypic variability and polyploidy in natural populations of the South American *Lathyrus nervosus* Lam. (Fabaceae). *Plant Syst Evol* **298**: 761–773. <http://link.springer.com/10.1007/s00606-011-0587-z> (Accessed December 9, 2013).
- Chao D-Y, Dilkes B, Luo H, Douglas A, Yakubova E, Lahner B, Salt DE. 2013. Polyploids exhibit higher potassium uptake and salinity tolerance in *Arabidopsis*. *Science* **341**: 658–9. <http://www.sciencemag.org/content/341/6146/658.short> (Accessed September 27, 2013).
- Le Comber SC, Ainouche ML, Kovarik A, Leitch AR. 2010. Making a functional diploid: from polysomic to disomic inheritance. *New Phytol* **186**: 113–22.

<http://www.ncbi.nlm.nih.gov/pubmed/20028473> (Accessed November 17, 2013).

Cornuet J-M, Ravigné V, Estoup A. 2010. Inference on population history and model checking using DNA sequence and microsatellite data with the software DIYABC (v1.0). *BMC Bioinformatics* **11**: 401. <http://dx.doi.org/10.1186/1471-2105-11-401>.

Coyne JA, Orr HA, others. 2004. *Speciation*. Sinauer Associates Sunderland, MA.

Csillery K, Francois O, Blum MGB. 2012. abc: an R package for approximate Bayesian computation (ABC). *Methods Ecol Evol*.

Doyle JJ, Egan AN. 2010. Dating the origins of polyploidy events. *New Phytol* **186**: 73–85. <http://www.ncbi.nlm.nih.gov/pubmed/20028472> (Accessed November 7, 2013).

Doyle JJ, Flagel LE, Paterson AH, Rapp RA, Soltis DE, Soltis PS, Wendel JF. 2008. Evolutionary genetics of genome merger and doubling in plants. *Annu Rev Genet* **42**: 443–61. <http://www.annualreviews.org/doi/abs/10.1146/annurev.genet.42.110807.091524> (Accessed November 12, 2013).

Durand EY, Patterson N, Reich D, Slatkin M. 2011. Testing for ancient admixture between closely related populations. *Mol Biol Evol* **28**: 2239–52. <http://mbe.oxfordjournals.org/content/28/8/2239.short> (Accessed August 6, 2013).

Fagundes NJR, Ray N, Beaumont M, Neuenschwander S, Salzano FM, Bonatto SL, Excoffier L. 2007. Statistical evaluation of alternative models of human evolution. *Proc Natl Acad Sci U S A* **104**: 17614–17619. <http://dx.doi.org/10.1073/pnas.0708280104>.

Fowler NL, Levin DA. 1984. Ecological constraints on the establishment of a novel polyploid in competition with its diploid progenitor. *Am Nat*. <http://agris.fao.org/agris-search/search.do?f=1985/US/US85101.xml;US8538711> (Accessed December 9, 2013).

François O, Blum MGB, Jakobsson M, Rosenberg NA. 2008. Demographic history of european populations of *Arabidopsis thaliana*. ed. T. Bataillon. *PLoS Genet* **4**: e1000075. <http://dx.plos.org/10.1371/journal.pgen.1000075> (Accessed August 7, 2013).

Gallardo MH, Kausel G, Jemenez A, Bacquet C, Gonzalez C, Figuereroa J, Kohler N, Ojeda R. 2004. Whole-genome duplications in South American desert rodents (Octodontidae). *Biol J Linn Soc* **82**: 443–451. <http://doi.wiley.com/10.1111/j.1095-8312.2004.00331.x> (Accessed September 11, 2013).

Gattepaille LM, Jakobsson M, Blum MG. 2013. Inferring population size changes with sequence and SNP data: lessons from human bottlenecks. *Heredity (Edinb)*. <http://dx.doi.org/10.1038/hdy.2012.120> (Accessed March 16, 2013).

Gayral P, Melo-Ferreira J, Glémin S, Bierne N, Carneiro M, Nabholz B, Lourenco JM, Alves PC, Ballenghien M, Faivre N, et al. 2013. Reference-free population genomics from next-generation transcriptome data and the vertebrate-invertebrate gap. ed. J.J. Welch. *PLoS Genet* **9**: e1003457. <http://dx.plos.org/10.1371/journal.pgen.1003457> (Accessed November 8, 2013).

Gregory RT, Mable BK. 2005. *The Evolution of the Genome*. Elsevier. Academic Press, San Diego <http://books.google.com/books?hl=en&lr=&id=8HtPZP9VSiMC&pgis=1> (Accessed September 11, 2013).

Gutiérrez JF, Vaquero F, Vences FJ. 1994. Allopolyploid vs. autopolyploid origins in the genus *Lathyrus* (Leguminosae)1. *Heredity (Edinb)* **73**: 29–40. <http://dx.doi.org/10.1038/hdy.1994.95> (Accessed November 16, 2013).

- Hamilton G, Currat M, Ray N, Heckel G, Beaumont M, Excoffier L. 2005. Bayesian estimation of recent migration rates after a spatial expansion. *Genetics* **170**: 409–417. <http://dx.doi.org/10.1534/genetics.104.034199>.
- Hardy OJ, Vanderhoeven S, De Loose M, Meerts P. 2000. Ecological, morphological and allozymic differentiation between diploid and tetraploid knapweeds (*Centaurea jacea*) from a contact zone in the Belgian Ardennes. *New Phytol* **146**: 281–290. <http://doi.wiley.com/10.1046/j.1469-8137.2000.00631.x> (Accessed November 17, 2013).
- Harrison RG. 1993. *Hybrid zones and the evolutionary process*. Oxford University Press <http://books.google.com/books?hl=en&lr=&id=aFJfKVKskYIC&pgis=1> (Accessed May 21, 2013).
- Hudson RR. 2002. Generating samples under a Wright-Fisher neutral model of genetic variation. *Bioinformatics* **18**: 337–338.
- Hudson RR. 1983. Properties of a neutral allele model with intragenic recombination. *Theor Popul Biol* **23**: 183–201. [http://dx.doi.org/10.1016/0040-5809\(83\)90013-8](http://dx.doi.org/10.1016/0040-5809(83)90013-8) (Accessed September 12, 2013).
- Hurka H, Freundner S, Brown AHD, Plantholt U. 1989. Aspartate aminotransferase isozymes in the genus *Capsella* (Brassicaceae): Subcellular location, gene duplication, and polymorphism. *Biochem Genet* **27**: 77–90. <http://link.springer.com/10.1007/BF00563019> (Accessed January 20, 2014).
- Hurka H, Neuffer B. 1997. Evolutionary processes in the genus *Capsella* (Brassicaceae). *Plant Syst Evol* **206**: 295–316. <http://link.springer.com/10.1007/BF00987954> (Accessed January 20, 2014).
- Jakobsson M, Hagenblad J, Tavaré S, Säll T, Halldén C, Lind-Halldén C, Nordborg M. 2006. A unique recent origin of the allotetraploid species *Arabidopsis suecica*: Evidence from nuclear DNA markers. *Mol Biol Evol* **23**: 1217–31. <http://mbe.oxfordjournals.org/content/23/6/1217.short> (Accessed September 12, 2013).
- Jones G, Sagitov S, Oxelman B. 2013. Statistical inference of allopolyploid species networks in the presence of incomplete lineage sorting. *Syst Biol* **62**: 467–78. <http://sysbio.oxfordjournals.org/content/62/3/467.short> (Accessed August 21, 2013).
- Justin R, W. SD. 2002. Annual Review of Ecology and Systematics. <http://archive.neicon.ru/xmlui/handle/123456789/1927630> (Accessed September 27, 2013).
- Kamiri M, Stift M, Srairi I, Costantino G, El Moussadik A, Hmyene A, Bakry F, Ollitrault P, Froelicher Y. 2011. Evidence for non-disomic inheritance in a Citrus interspecific tetraploid somatic hybrid between *C. reticulata* and *C. limon* using SSR markers and cytogenetic analysis. *Plant Cell Rep* **30**: 1415–25. <http://www.ncbi.nlm.nih.gov/pubmed/21409551> (Accessed December 9, 2013).
- Koning-Boucoiran CFS, Gitonga VW, Yan Z, Dolstra O, van der Linden CG, van der Schoot J, Uenk GE, Verlinden K, Smulders MJM, Krens FA, et al. 2012. The mode of inheritance in tetraploid cut roses. *Theor Appl Genet* **125**: 591–607. <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3397129&tool=pmcentrez&rendertype=abstract> (Accessed November 11, 2013).
- Lahr MM. 1994. The Multiregional Model of modern human origins: a reassessment of its morphological basis. *J Hum Evol* **26**: 23–56. <http://dx.doi.org/10.1006/jhev.1994.1003>

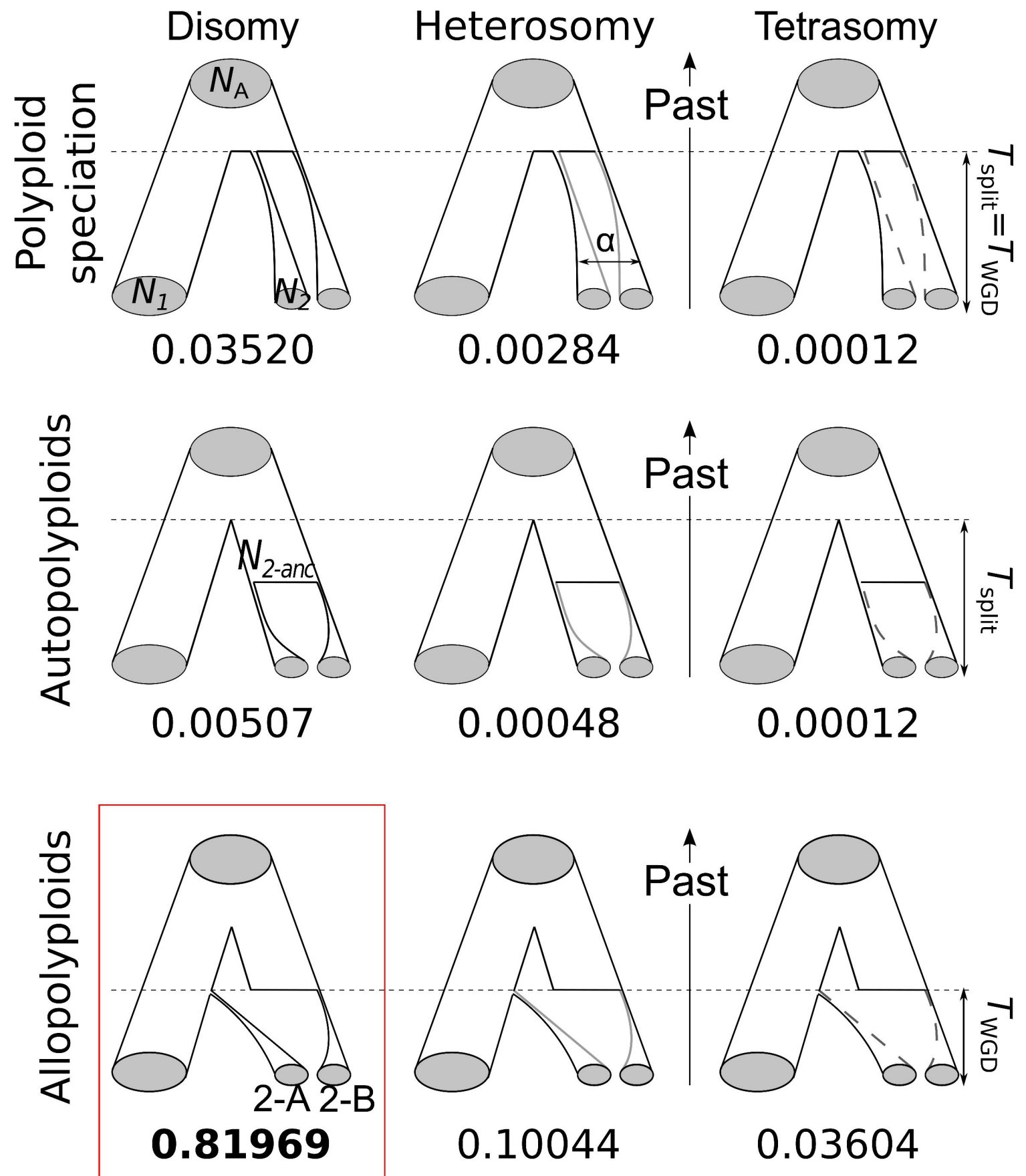
(Accessed September 11, 2013).

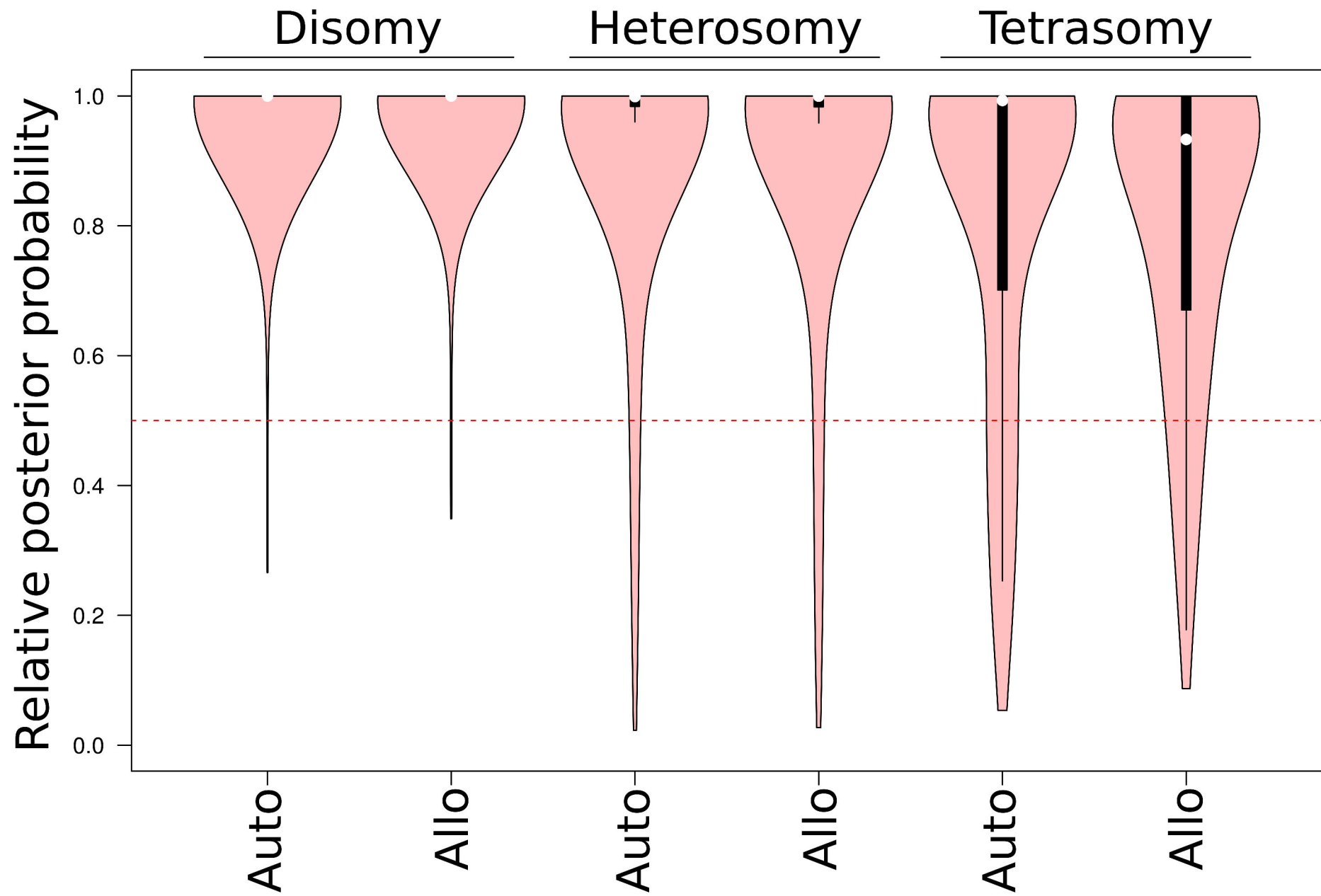
- Laval G, Excoffier L. 2004. SIMCOAL 2.0: a program to simulate genomic diversity over large recombining regions in a subdivided population with a complex history. *Bioinformatics* **20**: 2485–7. <http://bioinformatics.oxfordjournals.org/content/20/15/2485.short> (Accessed August 13, 2013).
- Lepais O, Muller SD, Ben Saad-Limam S, Benslama M, Rhazi L, Belouahem-Abed D, Daoud-Bouattour A, Gammar AM, Ghrabi-Gammar Z, Bacles CFE. 2013. High genetic diversity and distinctiveness of rear-edge climate relicts maintained by ancient tetraploidisation for *Alnus glutinosa*. ed. S. Lavergne. *PLoS One* **8**: e75029. <http://dx.plos.org/10.1371/journal.pone.0075029> (Accessed January 22, 2014).
- Levin DA. 1983. Polyploidy and novelty in flowering plants. *Am Nat* **1**: 1–25.
- Li H, Stephan W. 2006. Inferring the demographic history and rate of adaptive substitution in *Drosophila*. ed. M. Przeworski. *PLoS Genet* **2**: e166. <http://dx.plos.org/10.1371/journal.pgen.0020166> (Accessed August 13, 2013).
- Linder CR, Rieseberg LH. 2004. Reconstructing patterns of reticulate evolution in plants. *Am J Bot* **91**: 1700–8. <http://www.amjbot.org/content/91/10/1700.short> (Accessed August 8, 2013).
- Lynch M, Conery J. 2000. The Evolutionary Fate and Consequences of Duplicate Genes. *Science (80-)* **290**: 1151–1155. <http://www.sciencemag.org/content/290/5494/1151.short> (Accessed November 7, 2013).
- Mallet J. 2007. Hybrid speciation. *Nature* **446**: 279–283. <http://dx.doi.org/10.1038/nature05706>.
- Mayrose I, Zhan SH, Rothfels CJ, Magnuson-Ford K, Barker MS, Rieseberg LH, Otto SP. 2011. Recently formed polyploid plants diversify at lower rates. *Science* **333**: 1257. <http://www.sciencemag.org/content/333/6047/1257.abstract> (Accessed September 7, 2013).
- Meirmans PG, Van Tienderen PH. 2013. The effects of inheritance in tetraploids on genetic diversity and population divergence. *Heredity (Edinb)* **110**: 131–7. <http://www.ncbi.nlm.nih.gov/pubmed/23211786> (Accessed August 20, 2013).
- Mellars P, Stringer C. 1989. *The Human Revolution: Behavioural and Biological Perspectives on the Origins of Modern Humans*. http://books.google.ch/books/about/The_Human_Revolution.html?id=oWUUnAQAAIAAJ&pgis=1 (Accessed September 11, 2013).
- Mráz P, Garcia-Jacas N, Gex-Fabry E, Susanna A, Barres L, Müller-Schärer H. 2012. Allopolyploid origin of highly invasive *Centaurea stoebe* s.l. (Asteraceae). *Mol Phylogenet Evol* **62**: 612–623. <http://www.sciencedirect.com/science/article/pii/S1055790311004775> (Accessed November 16, 2013).
- Neiman M, Paczesniak D, Soper DM, Baldwin AT, Hehman G. 2011. Wide variation in ploidy level and genome size in a New Zealand freshwater snail with coexisting sexual and asexual lineages. *Evolution* **65**: 3202–16. <http://www.ncbi.nlm.nih.gov/pubmed/22023586> (Accessed September 11, 2013).
- Nielsen R. 2005. Molecular signatures of natural selection. *Annu Rev Genet* **39**: 197–218. <http://www.annualreviews.org/doi/full/10.1146/annurev.genet.39.073003.112420?prevSearch=Molecular+signatures+of+natural+selection&searchHistoryKey=> (Accessed August 6, 2013).

- Obbard DJ, Harris SA, Buggs RJA, Pannell JR. 2006. Hybridization, Polyploidy, And The Evolution Of Sexual Systems In *Mercurialis* (Euphorbiaceae). *Evolution (N Y)* **60**: 1801–1815. <http://doi.wiley.com/10.1111/j.0014-3820.2006.tb00524.x> (Accessed September 27, 2013).
- Olson MS. 1997. Bayesian Procedures for Discriminating Among Hypotheses With Discrete Distributions: Inheritance in the Tetraploid *Astilbe biternata*. *Genetics* **147**: 1933–1942. <http://www.genetics.org/content/147/4/1933.short> (Accessed November 13, 2013).
- Otto SP. 2007. The Evolutionary Consequences of Polyploidy. *Cell* **131**: 452–462. <http://www.sciencedirect.com/science/article/pii/S0092867407013402> (Accessed December 9, 2013).
- Otto SP, Whitton J. 2000. Polyploid incidence and evolution. *Annu Rev Genet* **34**: 401–437. <http://www.ncbi.nlm.nih.gov/pubmed/11092833> (Accessed August 13, 2013).
- Pannell JR, Obbard DJ, Buggs RJA. 2004. Polyploidy and the sexual system: what can we learn from *Mercurialis annua*? *Biol J Linn Soc* **82**: 547–560. <http://doi.wiley.com/10.1111/j.1095-8312.2004.00340.x> (Accessed January 28, 2014).
- Parisod C, Holderegger R, Brochmann C. 2010. Evolutionary consequences of autopolyploidy. *New Phytol* **186**: 5–17. <http://www.ncbi.nlm.nih.gov/pubmed/20070540> (Accessed November 7, 2013).
- Pritchard JK, Seielstad MT, Perez-Lezaun A, Feldman MW. 1999. Population growth of human Y chromosomes: a study of Y chromosome microsatellites. *Mol Biol Evol* **16**: 1791–1798.
- Pupilli F, Caceres ME, Arcioni S, Quarini CL. 1997. Segregation analysis of RFLP markers reveals a tetrasomic inheritance in apomictic *Paspalum simplex*. *Genome* **40**: 822–828. <http://www.nrcresearchpress.com/doi/abs/10.1139/g97-806> (Accessed December 9, 2013).
- Ramos-Onsins SE, Stranger BE, Mitchell-Olds T, Aguadé M. 2004. Multilocus analysis of variation and speciation in the closely related species *Arabidopsis halleri* and *A. lyrata*. *Genetics* **166**: 373–388.
- Reich DE, Cargill M, Bolk S, Ireland J, Sabeti PC, Richter DJ, Lavery T, Kouyoumjian R, Farhadian SF, Ward R, et al. 2001. Linkage disequilibrium in the human genome. *Nature* **411**: 199–204. <http://dx.doi.org/10.1038/35075590> (Accessed August 29, 2013).
- Rodriguez DJ. 1996. A Model For The Establishment Of Polyploidy In Plants. *Am Nat* **147**: 33–46. <http://cat.inist.fr/?aModele=afficheN&cpsidt=10889188> (Accessed December 9, 2013).
- Ross-Ibarra J, Tenaillon M, Gaut BS. 2009. Historical divergence and gene flow in the genus *Zea*. *Genetics* **181**: 1399–1413. <http://dx.doi.org/10.1534/genetics.108.097238>.
- Ross-Ibarra J, Wright SI, Foxe JP, Kawabe A, DeRose-Wilson L, Gos G, Charlesworth D, Gaut BS. 2008. Patterns of polymorphism and demographic history in natural populations of *Arabidopsis lyrata*. *PLoS One* **3**: e2411. <http://dx.doi.org/10.1371/journal.pone.0002411>.
- Roux C, Castric V, Pauwels M, Wright SI, Saumitou-Laprade P, Vekemans X. 2011. Does speciation between *Arabidopsis halleri* and *Arabidopsis lyrata* coincide with major changes in a molecular target of adaptation? *PLoS One* **6**: e26872. <http://dx.doi.org/10.1371/journal.pone.0026872>.
- Roux C, Pauwels M, Ruggiero M-V, Charlesworth D, Castric V, Vekemans X. 2012. Recent and Ancient Signature of Balancing Selection around the S-Locus in *Arabidopsis halleri* and *A. lyrata*. *Mol Biol Evol*. <http://dx.doi.org/10.1093/molbev/mss246>.
- Roux C, Tsagkogeorga G, Bierne N, Galtier N. 2013. Crossing the Species Barrier: Genomic

- Hotspots of Introgression between Two Highly Divergent *Ciona intestinalis* Species. *Mol Biol Evol* mst066-. <http://mbe.oxfordjournals.org/content/early/2013/04/05/molbev.mst066.abstract> (Accessed May 17, 2013).
- Salmon A, Udall JA, Jeddeloh JA, Wendel J. 2012. Targeted capture of homoeologous coding and noncoding sequence in polyploid cotton. *G3 (Bethesda)* **2**: 921–30. <http://www.g3journal.org/content/2/8/921.full> (Accessed September 12, 2013).
- Senchina DS, Alvarez I, Cronn RC, Liu B, Rong J, Noyes RD, Paterson AH, Wing RA, Wilkins TA, Wendel JF. 2003. Rate variation among nuclear genes and the age of polyploidy in *Gossypium*. *Mol Biol Evol* **20**: 633–43. <http://mbe.oxfordjournals.org/content/20/4/633.short> (Accessed November 18, 2013).
- Slotte T, Ceplitis A, Neuffer B, Hurka H, Lascoux M. 2006. Intrageneric phylogeny of *Capsella* (Brassicaceae) and the origin of the tetraploid *C. bursa-pastoris* based on chloroplast and nuclear DNA sequences. *Am J Bot* **93**: 1714–24. <http://www.amjbot.org/content/93/11/1714.short> (Accessed January 20, 2014).
- Slotte T, Huang H, Lascoux M, Ceplitis A. 2008. Polyploid speciation did not confer instant reproductive isolation in *Capsella* (Brassicaceae). *Mol Biol Evol* **25**: 1472–81. <http://mbe.oxfordjournals.org/content/25/7/1472.short> (Accessed September 12, 2013).
- Soltis DE, Buggs RJA, Doyle JJ, Soltis PS. 2010. What we still don't know about polyploidy. <http://www.ingentaconnect.com/content/iapt/tax/2010/00000059/00000005/art00006> (Accessed December 9, 2013).
- Soltis DE, Soltis PS, Tate JA. 2003. Advances in the study of polyploidy since Plant speciation. *New Phytol* **161**: 173–191. <http://doi.wiley.com/10.1046/j.1469-8137.2003.00948.x> (Accessed September 11, 2013).
- St Onge KR, Foxe JP, Li J, Li H, Holm K, Corcoran P, Slotte T, Lascoux M, Wright SI. 2012. Coalescent-based analysis distinguishes between allo- and autopolyploid origin in Shepherd's Purse (*Capsella bursa-pastoris*). *Mol Biol Evol* **29**: 1721–33. <http://mbe.oxfordjournals.org/content/29/7/1721.short> (Accessed August 21, 2013).
- Stift M, Berenos C, Kuperus P, van Tienderen PH. 2008. Segregation models for disomic, tetrasomic and intermediate inheritance in tetraploids: a general procedure applied to Rorippa (yellow cress) microsatellite data. *Genetics* **179**: 2113–23. <http://www.genetics.org/content/179/4/2113.short> (Accessed September 16, 2013).
- Tajima F. 1983. Evolutionary relationship of DNA sequences in finite populations. *Genetics* **105**: 437–460.
- Tajima F. 1989. The effect of change in population size on DNA polymorphism. *Genetics* **123**: 597–601.
- Tavaré S, Balding DJ, Griffiths RC, Donnelly P. 1997. Inferring coalescence times from DNA sequence data. *Genetics* **145**: 505–518.
- Templeton A. 2006. *Population Genetics and Microevolutionary Theory*. Wiley-Liss.
- Tymowska J, Fischberg M. 1982. A comparison of the karyotype, constitutive heterochromatin, and nucleolar organizer regions of the new tetraploid species *Xenopus epitropicalis* Fischberg and Picard with those of *Xenopus tropicalis* Gray (Anura, Pipidae). *Cytogenet Genome Res* **34**: 149–157. <http://www.karger.com/Article/FullText/131803> (Accessed December 9, 2013).

- Wakeley J. 2009. *Coalescent theory: an introduction*. Roberts & Company Publishers.
- Wakeley J, Hey J. 1997. Estimating ancestral population parameters. *Genetics* **145**: 847–855.
- Wang J. 2004. Application of the One-Migrant-per-Generation Rule to Conservation and Management. *Conserv Biol* **18**: 332–343. <http://doi.wiley.com/10.1111/j.1523-1739.2004.00440.x> (Accessed September 15, 2013).
- Wang W-K, Ho C-W, Hung K-H, Wang K-H, Huang C-C, Araki H, Hwang C-C, Hsu T-W, Osada N, Chiang T-Y. 2010. Multilocus analysis of genetic divergence between outcrossing *Arabidopsis* species: evidence of genome-wide admixture. *New Phytol* **188**: 488–500. <http://www.ncbi.nlm.nih.gov/pubmed/20673288> (Accessed September 12, 2013).
- Watterson G. 1975. On the number of segregating sites in genetical models without recombination. *Theor Popul Biol* **7**: 256–276.
- Whitlock MC, McCauley DE. 1999. Indirect measures of gene flow and migration: $F_{ST} \approx 1/(4Nm+1)$. *Heredity (Edinb)* **82**: 117–125. <http://dx.doi.org/10.1038/sj.hdy.6884960> (Accessed September 15, 2013).
- Wood TE, Takebayashi N, Barker MS, Mayrose I, Greenspoon PB, Rieseberg LH. 2009. The frequency of polyploid speciation in vascular plants. *Proc Natl Acad Sci U S A* **106**: 13875–9. <http://www.pnas.org/content/106/33/13875.short> (Accessed September 11, 2013).





$T_{\text{split}} = T_{\text{WGD}}$

Autopolyploidy

Allopolyploidy

Relative posterior probability

1.0
0.8
0.6
0.4
0.2
0.0

Diso

Hetero

Tetra

Diso

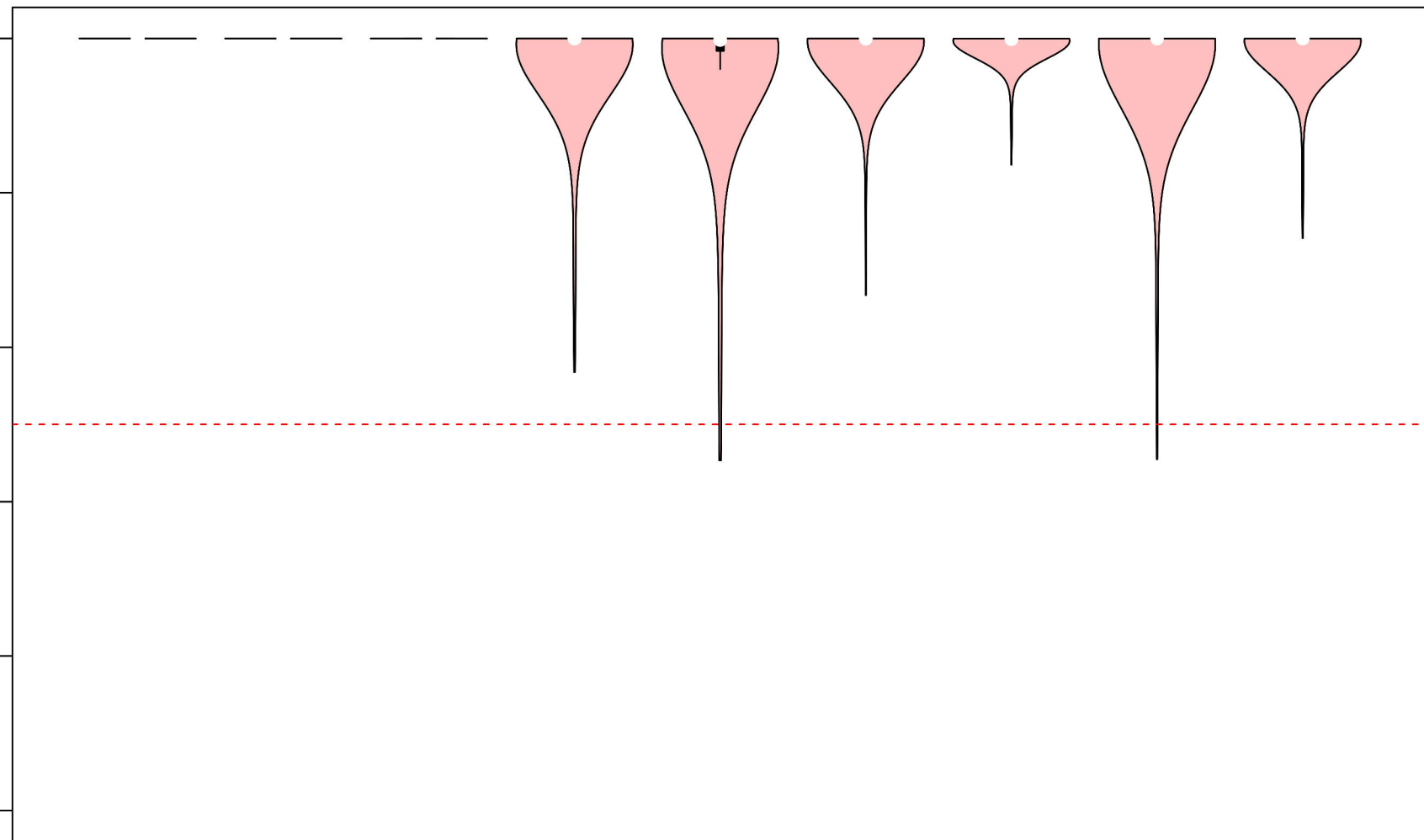
Hetero

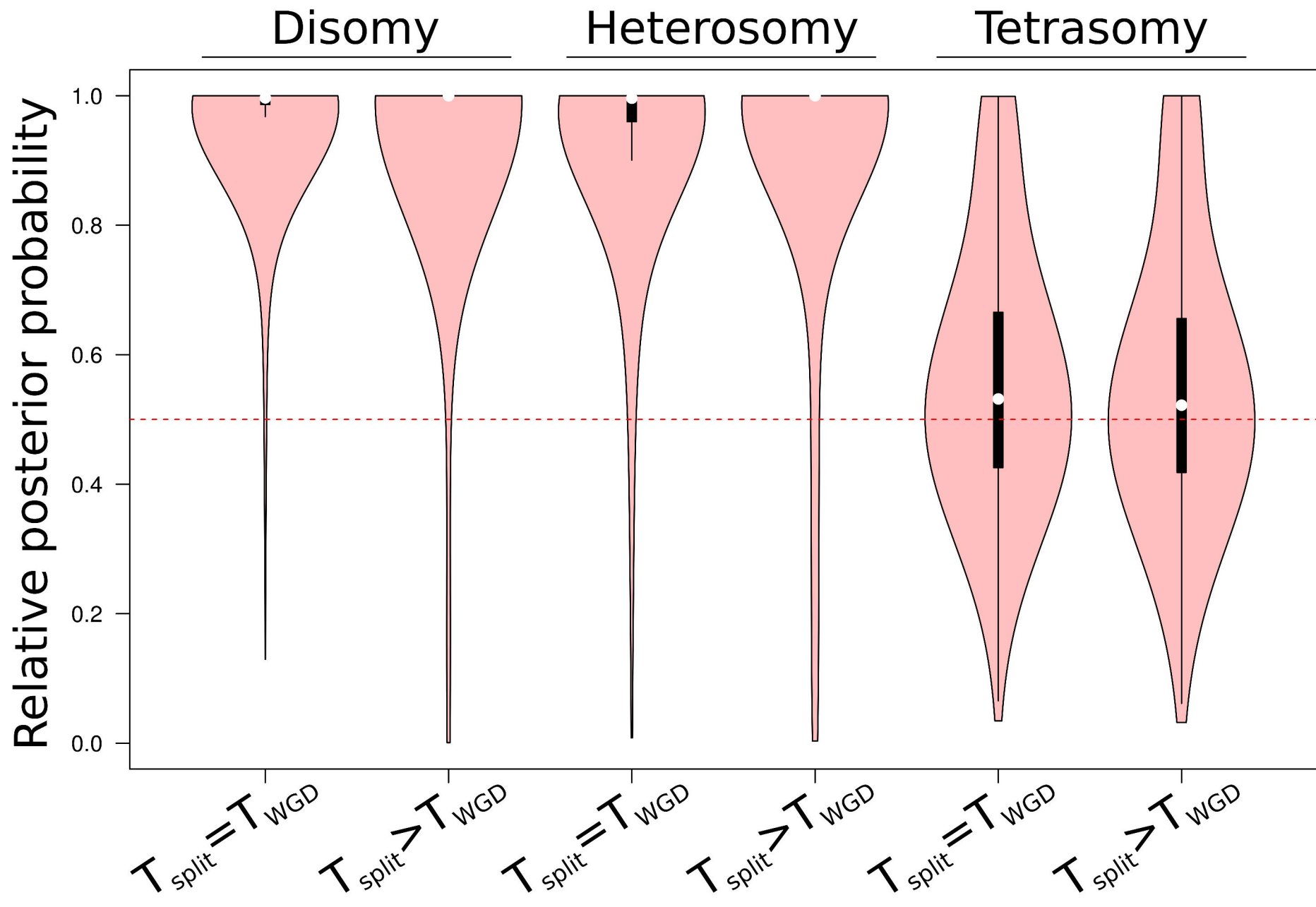
Tetra

Diso

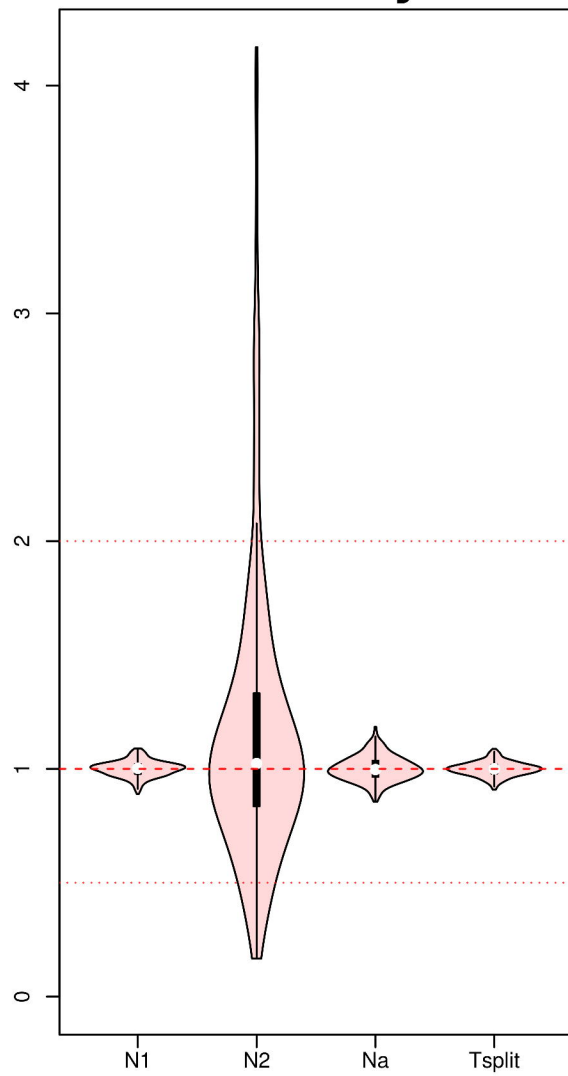
Hetero

Tetra

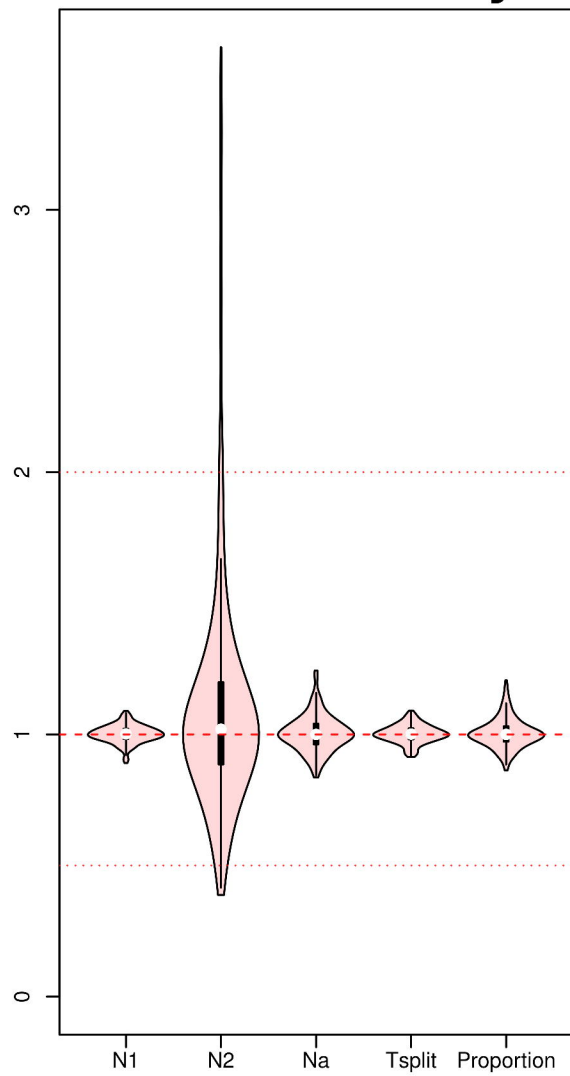




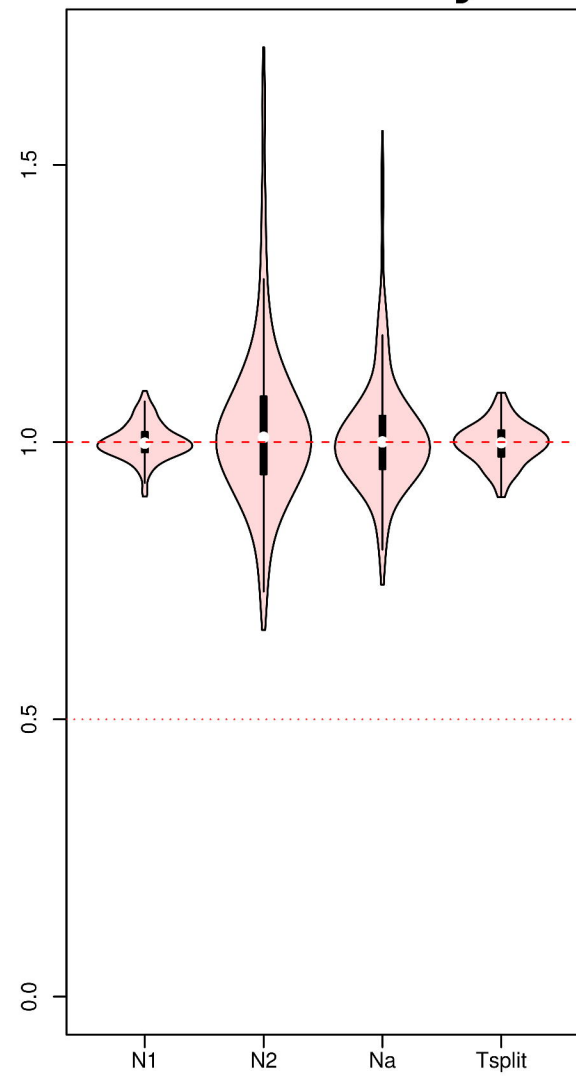
Disomy



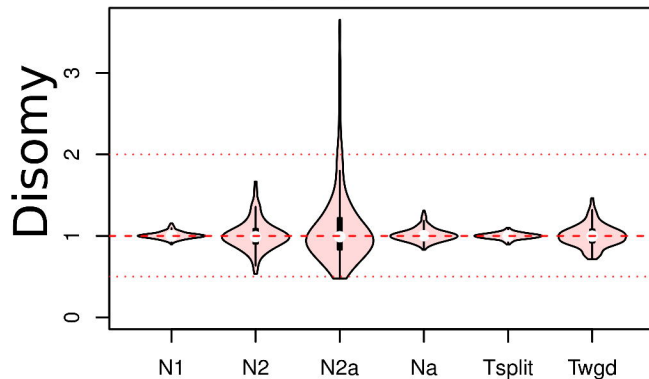
Heterosomy



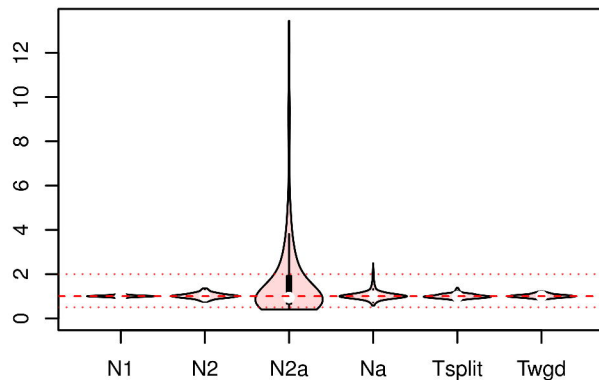
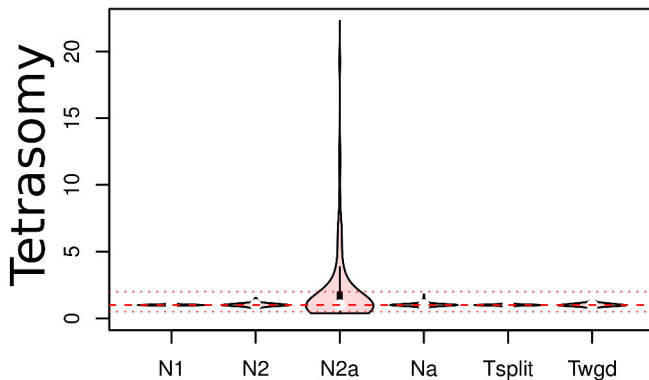
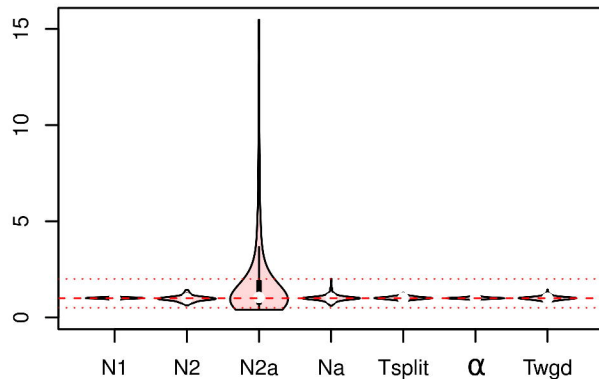
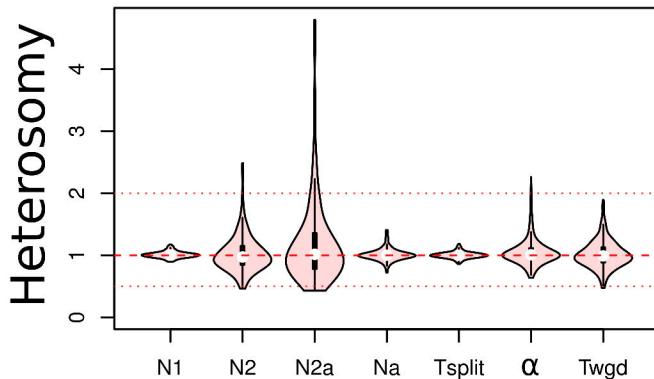
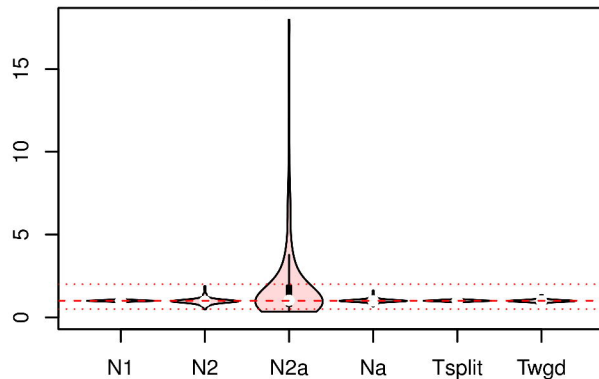
Tetrasomy

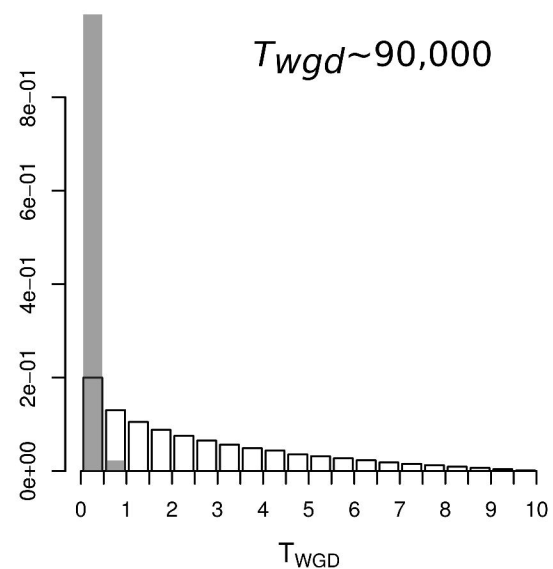
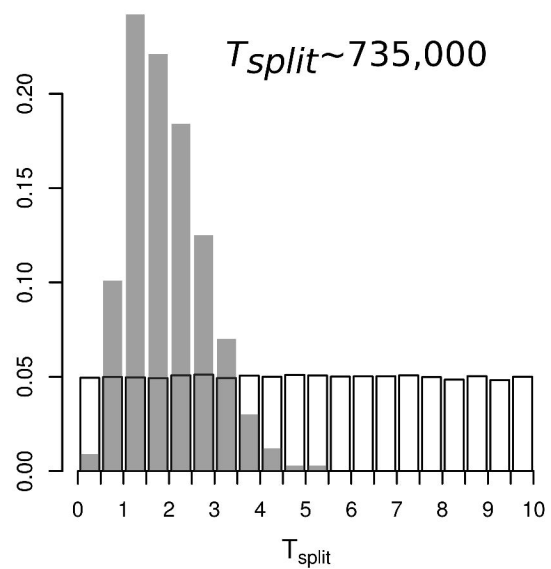
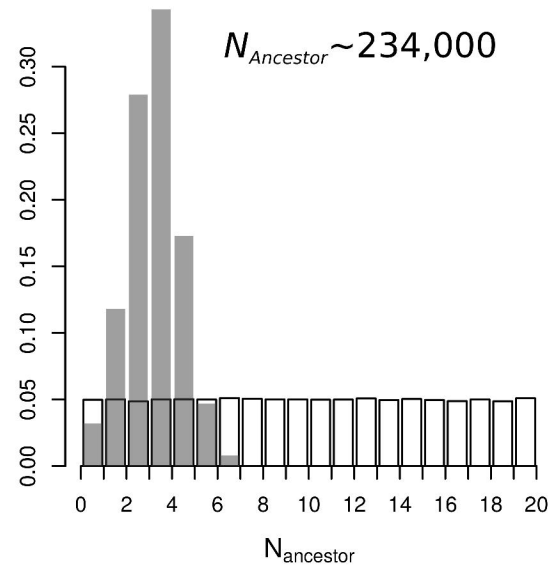
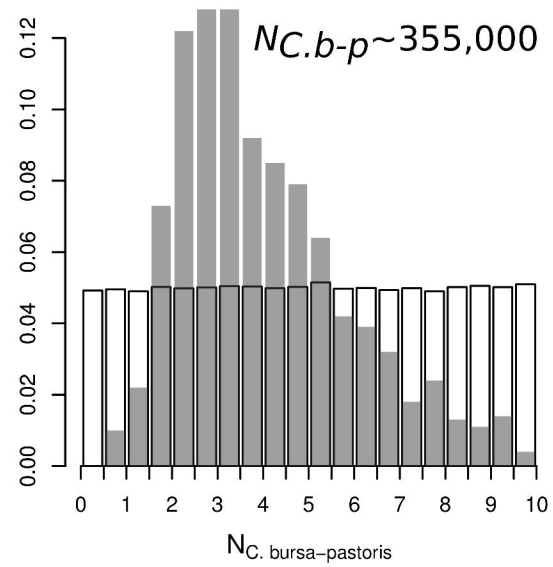
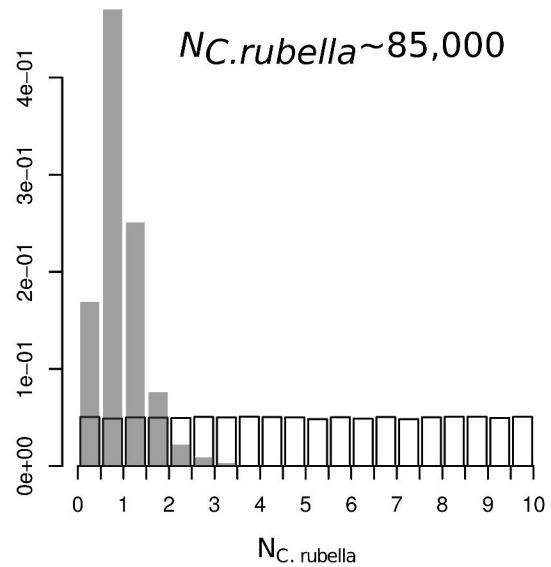


Autopolyploidization



Allopolyploidization





□ Prior
■ Posterior